IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Patent Application

Applicant(s):	Chen et al.
Docket No.:	YO999-172
Serial No.:	09/345,238
Filing Date:	June 30, 1999
Group:	2654
Examiner:	Qi Han

I hereby certify that this paper is being deposited on this date with the U.S. Postal Service as first class mail addressed to the Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

Signature: Tina Maurice Date: October 3, 2003

Title: Method and Apparatus for Tracking Speakers in an Audio Stream

APPEAL BRIEF

Mail Stop Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

Applicants hereby appeal the final rejection dated March 7, 2003, of claims 1 through 35 of the above-identified patent application.

REAL PARTY IN INTEREST

The present application is assigned to IBM Corporation, as evidenced by an assignment recorded on September 28, 1999 in the United States Patent and Trademark Office at Reel 010271, Frame 0023. The assignee, IBM Corporation, is the real party in interest.

RELATED APPEALS AND INTERFERENCES

There are no related appeals or interferences.

STATUS OF CLAIMS

Claims 1 through 35 are pending in the above-identified patent application. Claims 1-5, 8, 10-14, 16-19, 21-26 and 28-35 remain rejected under 35 U.S.C. §102(b) as being anticipated by Chen et al., "Speaker, Environment and Channel

Change Detection and Cluster via the Bayesian Information Criterion," Proc. of the DARPA Broadcast News Workshop (Feb. 1998), hereinafter, referred to as "Chen." In addition, Claims 6 and 7 remain rejected under 35 U.S.C. §103(a) as being unpatentable over Chen in view of well known prior art and claim 15 remains rejected under 35 U.S.C. §103(a) as being unpatentable over Chen in view of Kleider et al. (United States Patent No. 5,157,763).

STATUS OF AMENDMENTS

There have been no amendments filed subsequent to the final rejection.

SUMMARY OF INVENTION

The present invention automatically identifies speakers in an audio source by concurrently segmenting the audio source and clustering the segments corresponding to the same speaker (page 7, line 10, to page 8, line 14; page 13, line 10, to page 16, line 14).

ISSUES PRESENTED FOR REVIEW

i. Whether claims 1-5, 8, 10-14, 16-19, 21-26 and 28-35 are properly rejected under 35 U.S.C. §102(b) as being anticipated by Chen et al., "Speaker, Environment and Channel Change Detection and Cluster via the Bayesian Information Criterion," Proc. of the DARPA Broadcast News Workshop (Feb. 1998), hereinafter, referred to as "Chen";

ii. Whether claims 6 and 7 are properly rejected under 35 U.S.C. §103(a) as being unpatentable over Chen in view of well known prior art; and

iii. Whether claim 15 is properly rejected under 35 U.S.C. §103(a) as being unpatentable over Chen in view of Kleider et al.

GROUPING OF CLAIMS

The rejected claims stand and fall together.

ARGUMENT

Independent claims 1, 16, 23 and 30-35 are rejected under 35 U.S.C. §102(b) as being anticipated by Chen et al.

In the Office Action dated August 27, 2002, the Examiner asserted that Chen discloses speaker, environment and channel change detection and clustering via the Bayesian Information Criterion for segmenting the audio stream into homogeneous regions according to speaker identity, environmental condition and channel condition and clustering speech segments into homogeneous clusters according to speaker identity, environmental condition and channel (citing page 1, paragraph 2) which reads on the claimed "method of tracking a speaker in an audio source, said method comprising the steps of identifying potential segment boundaries in said audio source; and clustering homogeneous segments from said audio source substantially concurrently with said identifying step."

In the Response to Office Action dated December 26, 2002, Applicants submitted that while Chen does disclose segmenting an audio stream into homogeneous regions and clustering speech segments into homogeneous clusters, the audio stream is *first* segmented and *then* clustered. Applicants noted, as further evidence that the clustering in Chen is performed only after the audio stream has been segmented, that Section 4.1 indicates that each segment is compared to all other segments before clustering is finalized. In addition, Section 4.2, first paragraph indicates that the data set consists of an audio file that has been "hand-segmented into 824 short segments."

In the present Office Action, the Examiner notes that the prior art cites that "our segmentation algorithm can successfully detect acoustic changes" (Chen: abstract) and that "we first examine whether our detected change points were true." (Chen: Section 3.3, paragraph 3.) The Examiner asserts that this suggests that Chen not only employs its own segmenting mechanism, but is also capable of combining segmentation with clustering "substantially concurrently."

The Examiner also asserts that Chen suggests that clustering does not need completely segmented data, such that a clustering process may be combined with a segmenting process together substantially concurrently, since Chen discloses that "it is

also clear that our criterion can be applied to top-down methods." (Chen: Section 4.1, paragraph 4.)

The Examiner further asserts that a clustering step can be inserted in the segmentation loop, in Chen, Section 3.2, paragraph 1, and that Chen is capable of combining segmentation and clustering since the segmentation and clustering algorithms are based on the BIC algorithm and since equations (2), (3), and (8) have no limitation for combining segmentation and clustering.

Applicants acknowledge that Chen employs its own segmenting mechanism, but find no indication of or suggestion to perform segmentation and clustering "substantially concurrently" in the cited text. Applicants note that the Examiner asserts that Chen is *capable of* this, but does not assert that Chen suggests or discloses combining segmentation with clustering substantially concurrently.

Applicants note that, in the top-down method, a hypothesis is made regarding the number of clusters. Then, a test is made to determine if the number of clusters hypothesized actually "fits" the data. Alternatively, in the bottom-up method, the number of clusters is determined from the data. Thus, the capability to utilize a top-down method does not suggest that segmentation is performed substantially concurrently with the clustering process.

Regarding the final assertion made by the Examiner, Applicants also note that, whether or not Chen is capable of combining segmentation and clustering, there is no disclosure or suggestion to do so.

Thus, Chen does not disclose or suggest a "method of tracking a speaker in an audio source, said method comprising the steps of identifying potential segment boundaries in said audio source; and clustering homogeneous segments from said audio source substantially concurrently with said identifying step," as required by independent claims 1, 16, 30, 31, 32 and 33 of the present invention. Similarly, independent claims 23, 34 and 35 require that the segmentation and clustering are performed on the "same pass" through said audio source.

Additional Cited References

Kleider et al. was also cited by the Examiner in rejecting claims 15 for its disclosure that the information of the speaker model data may include a speaker name.

Applicants note that the inventors listed in United States Patent Number 5,157,763 (referred to by the Examiner in the Final Ofiice Action) are not Kleider et al. Applicants did find, however, United States Patent Number 5,930,748 in the Notice of References Cited and respond to that reference below.

Applicants note that Kleider et al. is directed to a "method of identifying an individual from a predetermined set of individuals using a speech sample spoken by the individual. The speech sample comprises a plurality of spoken utterance, and each individual of the set has predetermined speaker model data." Cited, Summary of the Invention. Kleider et al. do not address the issue of segmenting speech.

Thus, Kleider et al. do not disclose or suggest a "method of tracking a speaker in an audio source, said method comprising the steps of identifying potential segment boundaries in said audio source; and clustering homogeneous segments from said audio source substantially concurrently with said identifying step," as required by independent claims 1, 16, 30, 31, 32 and 33 of the present invention. Similarly, independent claims 23, 34 and 35 require that the segmentation and clustering are performed on the "same pass" through said audio source.

Conclusion

The rejections of the independent claims under section §103 in view of Chen, Kleider et al. or well known prior art, alone or in any combination, are therefore believed to be improper and should be withdrawn. The rejected dependent claims are believed allowable for at least the reasons identified above with respect to the independent claims.

The attention of the Examiner and the Appeal Board to this matter is appreciated.

Respectfully,

Kevin M. Mason

Date: October 3, 2003

Kevin M. Mason
Attorney for Applicant(s)
Reg. No. 36,597
Ryan, Mason & Lewis, LLP
1300 Post Road, Suite 205
Fairfield, CT 06824
(203) 255-6560

APPENDIX

1. A method for tracking a speaker in an audio source, said method comprising the steps of:

identifying potential segment boundaries in said audio source; and

clustering homogeneous segments from said audio source substantially concurrently with said identifying step.

2. The method of claim 1, wherein said identifying step identifies segment boundaries using a BIC model-selection criterion.

3. The method of claim 2, wherein a first model assumes there is no boundary in a portion of said audio source and a second model assumes there is a boundary in said portion of said audio source.

4. The method of claim 2, wherein a given sample, i, in said audio source is likely to be segment boundary if the following expression is negative:

$$\Delta BIC_i = -\frac{n}{2}\log|\Sigma_w| + \frac{i}{2}\log|\Sigma_f| + \frac{n-i}{2}\log|\Sigma_s| + \frac{1}{2}\lambda\left(d + \frac{d(d+1)}{2}\right)\log n$$

where $|\Sigma_w|$ is the determinant of the covariance of the window of all n samples, $|\Sigma_f|$ is the determinant of the covariance of the first subdivision of the window, and $|\Sigma_s|$ is the determinant of the covariance of the second subdivision of the window.

5. The method of claim 1, wherein said identifying step considers a smaller window size, n, of samples in areas where a segment boundary is unlikely to occur.

6. The method of claim 5, wherein said window size, n, is increased in a relatively slow manner when the window size is small and increases in a faster manner when the window size is larger.

7. The method of claim 5, wherein said window size, n, is initialized to a minimum value after a segment boundary is detected.

8. The method of claim 2, wherein said BIC model selection test is not performed at the border of each window of samples.

9. The method of claim 2, wherein said BIC model selection test is not performed when the window size, n, exceeds a predefined threshold.

10. The method of claim 1, wherein said clustering step is performed using a BIC model-selection criterion.

11. The method of claim 10, wherein a first model assumes that two segments or clusters should be merged, and a second model assumes that said two segments or clusters should be maintained independently.

12. The method of claim 11, further comprising the step of merging said two clusters if a difference in BIC values for each of said models is positive.

13. The method of claim 1, wherein said clustering step is performed using K previously identified clusters and M segments to be clustered.

14. The method of claim 1, further comprising the step of assigning a cluster identifier to each of said clusters.

15. The method of claim 1, further comprising the step of processing said audio source with a speaker identification engine to assign a speaker name to each of said clusters.

16. A method for tracking a speaker in an audio source, said method comprising the steps of:

identifying potential segment boundaries in said audio source; and

clustering segments from said audio source corresponding to the same speaker substantially concurrently with said identifying step.

17. The method of claim 16, wherein said identifying step identifies segment boundaries using a BIC model-selection criterion.

18. The method of claim 17, wherein a first model assumes there is no boundary in a portion of said audio source and a second model assumes there is a boundary in said portion of said audio source.

19. The method of claim 16, wherein said identifying step considers a smaller window size, n, of samples in areas where a segment boundary is unlikely to occur.

20. The method of claim 17, wherein said BIC model selection test is not performed where the detection of a boundary is unlikely to occur.

21. The method of claim 16, wherein said clustering step is performed using a BIC model-selection criterion, where a first model assumes that two segments or clusters should be merged, and a second model assumes that said two segments or clusters should be maintained independently.

22. The method of claim 16, wherein said clustering step is performed using K previously identified clusters and M segments to be clustered.

23. A method for tracking a speaker in an audio source, said method comprising the steps of:

identifying potential segment boundaries during a pass through said audio source; and

clustering segments from said audio source corresponding to the same speaker during said same pass through said audio source.

24. The method of claim 23, wherein said identifying step identifies segment boundaries using a BIC model-selection criterion.

25. The method of claim 24, wherein a first model assumes there is no boundary in a portion of said audio source and a second model assumes there is a boundary in said portion of said audio source.

26. The method of claim 23, wherein said identifying step considers a smaller window size, n, of samples in areas where a segment boundary is unlikely to occur.

27. The method of claim 24, wherein said BIC model selection test is not performed where the detection of a boundary is unlikely to occur.

28. The method of claim 23, wherein said clustering step is performed using a BIC model-selection criterion, where a first model assumes that two segments or clusters should be merged, and a second model assumes that said two segments or clusters should be maintained independently.

29. The method of claim 23, wherein said clustering step is performed using K previously identified clusters and M segments to be clustered.

30. A system for tracking a speaker in an audio source, comprising:

a memory that stores computer-readable code; and

a processor operatively coupled to said memory, said processor configured to implement said computer-readable code, said computer-readable code configured to:

identify potential segment boundaries in said audio source; and

cluster homogeneous segments from said audio source substantially concurrently with said identification of segment boundaries.

31. An article of manufacture, comprising:

a computer readable medium having computer readable code means embodied thereon, said computer readable program code means comprising:

a step to identify potential segment boundaries in said audio source; and

a step to cluster homogeneous segments from said audio source substantially concurrently with said identification of segment boundaries.

32. A system for tracking a speaker in an audio source, comprising:

a memory that stores computer-readable code; and

a processor operatively coupled to said memory, said processor configured to implement said computer-readable code, said computer-readable code configured to:

identify potential segment boundaries in said audio source; and

cluster segments from said audio source corresponding to the same speaker substantially concurrently with said identification of segment boundaries.

33. An article of manufacture, comprising:

a computer readable medium having computer readable code means embodied thereon, said computer readable program code means comprising:

a step to identify potential segment boundaries in said audio source; and

a step to cluster segments from said audio source corresponding to the same speaker substantially concurrently with said identification of segment boundaries.

34. A system for tracking a speaker in an audio source, comprising:

a memory that stores computer-readable code; and

a processor operatively coupled to said memory, said processor configured to implement said computer-readable code, said computer-readable code configured to:

identify potential segment boundaries during a pass through said audio source; and

cluster segments from said audio source corresponding to the same speaker during said same pass through said audio source.

35. An article of manufacture, comprising:

a computer readable medium having computer readable code means embodied thereon, said computer readable program code means comprising:

a step to identify potential segment boundaries during a pass through said audio source; and

a step to cluster segments from said audio source corresponding to the same speaker during said same pass through said audio source.